Available online at www.sciencedirect.com

ScienceDirect

Journal of Combinatorial Theory, Series A 113 (2006) 1746–1759

Journal of Combinatorial Theory Series A

www.elsevier.com/locate/jcta

Generic erasure correcting sets: Bounds and constructions

Henk D.L. Hollmann, Ludo M.G.M. Tolhuizen

Philips Research Laboratories, High Tech Campus HTC-36, 5656 AE Eindhoven, The Netherlands

Received 7 October 2005

Available online 21 June 2006

Abstract

A generic (r, m)-erasure correcting set generates for each binary linear code of codimension r a collection of parity check equations that enables iterative decoding of all potentially correctable erasure patterns of size at most m. As we have shown earlier, such a set essentially is just a parity check collection with this property for the Hamming code of codimension r.

We prove non-constructively that for fixed m the minimum size $F(r, m)$ of a generic (r, m)-erasure correcting set is linear in r. Moreover, we show constructively that $F(r, 3) \leqslant 3(r-1)^{\log_2 3} + 1$, which is a major improvement on a previous construction showing that $F(r, 3) \leqslant 1 + \frac{1}{2}r(r-1)$.

In the course of this work we encountered the following problem that may be of independent interest: what is the smallest size of a collection $C \subseteq \mathbb{F}_2^n$ such that, given any set of s independent vectors in $\mathbb{F}_2^n$, there is a vector $\mathbf{c} \in C$ that has inner product 1 with all of these vectors? We show non-constructively that, for fixed s, this number is linear in n.

Keywords: Binary erasure channel; Iterative decoding; Erasure decoding; Stopping set

1. Introduction

Consider the following well-known scheme for iterative decoding of a binary linear code C used on the binary erasure channel [1]. Fix some set $\mathcal{H}$ of parity check equations for C. Suppose that we receive a word with set of erased positions E, say. If one of the parity check equations

E-mail addresses: henk.d.l.hollmann@philips.com (H.D.L. Hollmann), ludo.tolhuizen@philips.com (L.M.G.M. Tolhuizen).

0097-3165/$ – see front matter
doi:10.1016/j.jcta.2006.03.024

from $\mathcal{H}$ involves *precisely one* of the erasures in E, then we use this equation to determine the value of this erasure, and we continue. If no such parity check equation can be found, then we stop; in that case, the set E is called a *stopping set for* $\mathcal{H}$ [1]. Since there is no way to resolve any of the erasures contained in a stopping set, a received word with set of erased positions E can be fully decoded precisely when E does not contain a non-empty stopping set. In this paper, when we speak of iterative decoding we always mean decoding by this scheme.

Note that (non-trivial) binary parity check equations are in one-to-one correspondence with (non-zero) codewords from the binary dual $C^{\perp}$ of C. For convenience, in what follows we will not distinguish between the two notions.

Using different sets $\mathcal{H}$ of parity check equations for C may result in different stopping sets. Note, however, that the support of a codeword is always a stopping set, since by definition each parity check vector has an even number of ones within such a set.

An erasure pattern is unambiguously decodable if there is only one way to fill in the erasures such that the resulting word is contained in C. As C is linear, this is the case if and only if the erasure pattern does not contain the support of a non-zero codeword. We therefore call an erasure pattern C-uncorrectable if it contains the support of a non-zero codeword, and C-correctable otherwise. We will speak more briefly of correctable and uncorrectable if it is clear form the context which code is referred to.

So no uncorrectable erasure pattern can be iteratively decoded; however, such a pattern cannot be (unambiguously) decoded by *any* algorithm. In [5] it has been shown that, conversely, all other erasure patterns (i.e., the correctable ones) can be iteratively decoded provided that the collection of parity checks $\mathcal{H}$ used is large enough. Here we will be interested in parity check collections that enable iterative decoding of all correctable erasure patterns of a sufficiently small size. Related work can be found in [2], where Weber and Abdel-Ghaffar construct collections of parity check equations for the Hamming code C_r of redundancy r that enable iterative decoding of all correctable erasure patterns of size at most three. Also related are [3] and [4], where Schwartz and Vardy study the minimum size of collections of parity check equations for a code C that enable iterative decoding of all erasure patterns of size less than the minimum distance of C (note that all such erasure patterns are C-correctable).

In [5], we introduced and constructed so-called *generic* (r, m)*-erasure correcting sets*. These are subsets $\mathcal{A}$ of $\mathbb{F}_2^r$ such that for *any* code C of length n and codimension r, and *any* $r \times n$ parity check matrix H for this code, the collection of parity check equations

$$\{\mathbf{a}H \mid \mathbf{a} \in \mathcal{A}\}$$

enables iterative decoding of all C-correctable erasure patterns of size at most m. Our aim is to construct generic (r, m)-erasure correcting sets of small size, and to investigate the minimum size $F(r, m)$ of such sets.

In Section 3 we will show non-constructively that for each $m \geqslant 1$, there exists a constant c_m such that $F(r, m) \leqslant c_m r$ for $r \geqslant m$. At present we do not have a *constructive* proof that $F(r, m)$ is linear in r.

From Section 4 onward, we only consider *special* generic (r, m)-erasure correcting sets, namely those that are contained in the complement of a hyperplane in $\mathbb{F}_2^r$ (i.e., in the complement of an $(r-1)$-dimensional subspace). First we characterize in two ways when a set contained in the complement of a hyperplane is generic (r, m)-erasure correcting. In Section 5, we use the first characterization to retrieve the explicit generic (r, m)-erasure correcting sets from [5], and to show that finding a special generic $(r, r-2)$-erasure correcting set of minimal size is equivalent to finding a longest code of codimension $r-1$ with minimum distance at least 5.

As shown in Section 6, the second characterization motivates the study of the following problem: given n and $s \leqslant n$, determine the smallest size $G(n, s)$ of a set $C \subseteq \mathbb{F}_2^n$ such that for any set of s independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_s$ in $\mathbb{F}_2^n$, there is a $\mathbf{c} \in C$ that has inner product 1 with all the vectors in the set. In Section 6.2, we give an explicit construction for the case $s = 2$, which is then used to show that $F(r, 3)$ is of order at most $r^{\log_2 3}$. Finally, in Section 6.3, we give upper and lower bounds on the size of special generic (r, m)-erasure sets. In particular, we show (again non-constructively) that $G(n, s)$ is linear in n, and use this result to show that for each $m \geqslant 3$, there exist special generic (r, m)-erasure correcting sets of size linear in r.

We finally remark that although this paper only considers the binary field, about all results here can easily be generalized to arbitrary finite fields.

2. Preliminaries

In this section, we introduce some notations and definitions. Throughout this paper, we use boldface letters to denote row vectors. All vectors and matrices are binary. If there is no confusion about the length of vectors, we denote by $\mathbf{0}$ and $\mathbf{1}$ the vectors consisting of only zeroes or only ones, and by $\mathbf{e}_i$ the ith unit vector, the vector that has a one in position i and zeroes elsewhere.

The size of a set A is denoted by $|A|$. If H is a $r \times n$ matrix and $E \subseteq \{1, 2, \ldots, n\}$, then the *restriction $H(E)$ of H to E* denotes the $r \times |E|$ matrix consisting of those columns of H indexed by E. Similarly, if $\mathbf{x} \in \mathbb{F}_2^n$ and $E \subseteq \{1, 2, \ldots, n\}$, then the restriction $\mathbf{x}(E)$ of $\mathbf{x}$ to E is the vector of length $|E|$ consisting of the entries indexed by E.

The *support* $\operatorname{supp}(\mathbf{x})$ of a vector $\mathbf{x} \in \mathbb{F}_2^n$ is the set of its non-zero coordinates, that is,

$$\operatorname{supp}(\mathbf{x}) = \{i \in \{1, 2, \ldots, n\} \mid x_i \neq 0\},$$

and the weight $\operatorname{wt}(\mathbf{x})$ of $\mathbf{x}$ is the size $|\operatorname{supp}(\mathbf{x})|$ of its support.

As usual, an $[n, k]$ code C is a k-dimensional subspace of $\mathbb{F}_2^n$; the dual code of C, denoted by $C^\perp$, is the $[n, r]$ code, $r = n - k$, consisting of all vectors in $\mathbb{F}_2^n$ that have inner product 0 with all words from C. The number r is referred to as the *codimension* or *redundancy* of the code. An $r \times n$ matrix is called a parity check matrix for C if its rows span $C^\perp$. When we speak about "code," we will always mean binary linear code.

The following definitions are taken from [5].

Definition 2.1. Let $C \subseteq \mathbb{F}_2^n$ be a code. A set $E \subseteq \{1, 2, \ldots, n\}$ is called *C-uncorrectable* if it contains the support of a non-zero codeword, and *C-correctable* otherwise.

The motivation for this definition is that a received word containing only correct symbols and erasures can be decoded unambiguously precisely when exactly one codeword agrees with this word in the non-erased positions; for linear codes this is the case precisely when the set of erasures does not contain the support of a non-zero codeword.

Definition 2.2. Let $C \subseteq \mathbb{F}_2^n$ be a code. A set $\mathcal{H} \subseteq C^\perp$ is called *m-erasure reducing for C* if for each erasure pattern $E \subseteq \{1, 2, \ldots, n\}$ of size m that is C-correctable, there exists a parity check equation $\mathbf{h}$ with exactly one 1 in the positions indexed by E, that is, with $\operatorname{wt}(\mathbf{h}(E)) = 1$. The set $\mathcal{H}$ is called *m-erasure correcting for C* if it is m'-erasure reducing for C for each m' with $1 \leqslant m' \leqslant m$.

In other words, if we use the iterative decoding algorithm with a set of parity check equations $\mathcal{H}$ that is m-erasure reducing for C, then from each C-correctable erasure pattern of size m at

least one erasure is resolved; if $\mathcal{H}$ is m-erasure correcting for C, then the iterative decoding algorithm in fact fully resolves all C-correctable erasure patterns of size *at most* m.

Remark. Note that if $\mathcal{H}$ is m-erasure reducing for a code C and if E is C-uncorrectable, but not equal to the support of a codeword, then an erasure from E may or may not be resolved. In other words, not every stopping set of size at most m necessarily is a codeword.

Finally, in [5] we introduced the notion of a "generic" m-erasure correcting and reducing set for codes of a fixed codimension. The idea is to describe which linear combinations to take given any parity check matrix for such a code.

Definition 2.3. Let $1 \leqslant m \leqslant r$. A set $\mathcal{A} \subseteq \mathbb{F}_2^r$ is called *generic* (r, m)*-erasure reducing* if for any $r \times n$ binary matrix H of rank r, the collection $\{\mathbf{a}H \mid \mathbf{a} \in \mathcal{A}\}$ is m-erasure reducing for the code with parity check matrix H.

Definition 2.4. Let $1 \leqslant m \leqslant r$. A set $\mathcal{A} \subseteq \mathbb{F}_2^r$ is called *generic* (r, m)*-erasure correcting* if it is generic (r, m')-erasure reducing for all m' with $1 \leqslant m' \leqslant m$.

In [5], it has been shown that for a particular code C, the notions "m-erasure reducing for C" and "m-erasure correcting for C" need not be the same; the notions "generic (r, m)-erasure reducing" and "generic (r, m)-erasure correcting," however, are in fact equivalent.

The following useful characterization of generic (r, m)-erasure correcting sets has been obtained in [5].

Proposition 2.5. *A set $\mathcal{A} \subseteq \mathbb{F}_2^r$ is generic (r, m)-erasure correcting if and only if for any $r \times m$ matrix M of rank m there is a vector $\mathbf{a} \in \mathcal{A}$ such that* $\mathrm{wt}(\mathbf{a}M) = 1$.

The proof of this proposition can be outlined as follows. It can be shown that an erasure pattern E is C-correctable if and only if for any parity check matrix H for C, the restriction $H(E)$ has full rank. Hence, we need only consider $r \times m$ submatrices of full rank, and each $r \times m$ matrix of full rank can occur as such a submatrix.

Note that any $r \times m$ matrix of rank m occurs, up to a column permutation, as a submatrix of any parity check matrix H_r of a $[2^r - 1, 2^r - r - 1]$ Hamming code C_r. As a consequence, a set $\mathcal{A} \subset \mathbb{F}_2^r$ is generic (r, m)-erasure correcting if and only if $\mathcal{H} = \{\mathbf{a}H_r \mid \mathbf{a} \in \mathcal{A}\}$ is m-erasure correcting for C_r.

We are interested in generic (r, m)-erasure correcting sets of small size. This motivates the following definition.

Definition 2.6. For $1 \leqslant m \leqslant r$, let $F(r, m)$ the smallest size of a generic (r, m)-erasure correcting set.

Note that Proposition 2.5 implies that $\mathbb{F}_2^r \setminus \{\mathbf{0}\}$ is generic (r, m)-erasure correcting, so $F(r, m)$ is well-defined.

3. Upper and lower bounds on $F(r, m)$

In this section, we will show that $F(r, m)$ is *of linear order* in r. To be precise, we will show that for each $m \geqslant 1$, there exists a constant c_m such that for each $r \geqslant m$, we have that $r \leqslant F(r, m) \leqslant c_m r$. Concerning the lower bound, we have the following lemma.

Lemma 3.1. *Any (r, m)-erasure correcting set spans $\mathbb{F}_2^r$. As a consequence, $F(r, m) \geqslant r$.*

Proof. (cf. [5]) Suppose $\mathcal{A} \subset \mathbb{F}_2^r$ is such that $\text{span}(\mathcal{A}) \neq \mathbb{F}_2^r$. We will show that $\mathcal{A}$ is not generic (r, m)-erasure correcting by constructing an $r \times m$ matrix M with rank m such that for each $\mathbf{a} \in \mathcal{A}$, the vector $\mathbf{a}M$ does not have weight 1 (cf. Proposition 2.5).

Let $\mathbf{v}$ be a non-zero vector that has inner product 0 with all words from $\mathcal{A}$. Let M be an invertible matrix such that the ith row of M has odd weight if and only if $i \in \text{supp}(\mathbf{v})$, and let $\mathbf{a} \in \mathcal{A}$. As $(\mathbf{v}, \mathbf{a}) = 0$, the vector $\mathbf{a}M$ is the sum of an even number of (odd weight) rows of M indexed by integers from $\text{supp}(\mathbf{v})$, and some (even weight) rows of M indexed by integers outside $\text{supp}(\mathbf{v})$. As a consequence, $\mathbf{a}M$ has even weight. □

The proof for the upper bound is non-constructive: we will show that the collection of all subsets of $\mathbb{F}_2^r$ of a sufficiently large size contains at least one generic (r, m)-erasure correcting set. The precise result is as follows.

Theorem 3.2. *For all $m \geqslant 1$ and $r \geqslant m$, we have that*

$$F(r, m) \leqslant \frac{m}{-\log_2(1 - m2^{-m})} \cdot r,$$

where $\log_2$ denotes the base-2 logarithm.

Proof. We give a probabilistic proof [6]. Let $1 \leqslant m \leqslant r$. We write $\mathcal{M}_{m,r}$ to denote the collection of all binary $r \times m$ matrices of rank m. Note that obviously

$$|\mathcal{M}_{m,r}| < 2^{rm}. \tag{1}$$

For any positive integer N, consider the following experiment. We construct a binary $N \times r$ matrix A by randomly setting each individual entry to zero or to one with probability $1/2$. Now interpret this matrix as a sequence of N row vectors $\mathbf{a}_1, \ldots, \mathbf{a}_N$, each of length r. For each matrix M in $\mathcal{M}_{m,r}$, we define the random variable X_M by

$$X_M = \begin{cases} 0, & \text{if there is an } i \text{ such that } \text{wt}(\mathbf{a}_i M) = 1; \\ 1, & \text{otherwise.} \end{cases}$$

So $X_M = 0$ if the matrix M is "good" with respect to the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_N$, and $X_M = 1$ if M is "bad." Furthermore, let the random variable X be defined as

$$X = \sum_{M \in \mathcal{M}_{m,r}} X_M.$$

Now X counts the number of bad matrices with respect to A; if $X < 1$, then $X = 0$ and all matrices are "good" with respect to A, hence the collection $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \subseteq \mathbb{F}_2^r$ satisfies the criterion in Proposition 2.5. Consequently, if $E[X] < 1$, then all matrices in $\mathcal{M}_{m,r}$ are good with respect to some matrix A, and so $F(r, m) \leqslant N$.

In order to compute $E[X]$, we will compute for each $M \in \mathcal{M}_{m,r}$ the probability $\mathrm{Prob}(X_M = 1)$ that X_M is equal to 1. Any such M has full rank, hence for each $i = 1, 2, \ldots, m$ there are exactly 2^{r-m} vectors $\mathbf{a}$ such that $\mathbf{a}M = \mathbf{e}_i$. We conclude that there are $m2^{r-m}$ "good" vectors for M, and hence $2^r(1 - m2^{-m})$ "bad" vectors. Now the matrix M is bad if all the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_N$ are bad; we conclude that

$$\mathrm{Prob}(X_M = 1) = \left(1 - m2^{-m}\right)^N.$$

Since expectation is a linear operation, we have that

$$E[X] = \sum_{M \in \mathcal{M}_{m,r}} E[X_M] = |\mathcal{M}_{m,r}|\left(1 - m2^{-m}\right)^N,$$

from which we conclude that $E[X] < 1$ if and only if

$$N > \frac{\log_2 |\mathcal{M}_{m,r}|}{-\log_2(1 - m2^{-m})}. \tag{2}$$

As a consequence of the foregoing, we have that $F(r, m) \leqslant N$ if N satisfies (2); by using (1) we see that this is certainly true if

$$N \geqslant \frac{m}{-\log_2(1 - m2^{-m})} r. \qquad \square$$

4. Special generic erasure-correcting sets: definition and characterizations

In the remainder of this paper, we only consider generic (r, m)-erasure correcting sets that are contained in the complement of a hyperplane in $\mathbb{F}_2^r$ (i.e., in the complement of an $(r-1)$-dimensional subspace). We call such sets *special* generic (r, m)-erasure correcting sets.

In this section, we will characterize in two ways when a set that is contained in the complement of a hyperplane actually is generic (r, m)-erasure correcting. Both characterizations will be used in subsequent sections. To this end, we first derive another characterization of generic (r, m)-erasure correcting sets.

Theorem 4.1. *A set $\mathcal{A} \subseteq \mathbb{F}_2^r$ is generic (r, m)-erasure correcting if and only if for any $(r - m)$-dimensional subspace U of $\mathbb{F}_2^r$ and for any m independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m$ in $\mathbb{F}_2^r$ with $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_m) \cap U = \{\mathbf{0}\}$, there is some $i \in \{1, \ldots, m\}$ for which $\mathcal{A}$ meets $\mathbf{b}_i + U$.*

Proof. There is a one-to-one correspondence between binary $r \times m$ matrices M of rank m and cosets $\mathbf{b}_i + U$ for $i = 1, \ldots, m$ as in the theorem. This correspondence is determined by the conditions that

$$U = \left\{\mathbf{x} \in \mathbb{F}_2^r \mid \mathbf{x}M = \mathbf{0}\right\} \quad \text{and} \quad \mathbf{b}_i M = \mathbf{e}_i \quad \text{for } i = 1, \ldots, m. \tag{3}$$

Indeed, first note that if U and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ satisfy (3), then $\mathbf{b}_1, \ldots, \mathbf{b}_m$ are independent and $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_m) \cap U = \{\mathbf{0}\}$; indeed, if $\mathbf{x} = \sum_i x_i \mathbf{b}_i \in U$, then

$$\mathbf{0} = \mathbf{x}M = \sum_i x_i \mathbf{e}_i,$$

hence $x_i = 0$ for all i, so that $\mathbf{x} = \mathbf{0}$. Note also that the conditions "$\mathrm{wt}(\mathbf{a}M) = 1$ for some $\mathbf{a} \in \mathcal{A}$" and "$\mathcal{A}$ meets some coset $\mathbf{b}_i + U$" are equivalent, since $\mathbf{b}_i + U = \{\mathbf{x} \in \mathbb{F}_2^r \mid \mathbf{x}M = \mathbf{e}_i\}$.

Conversely, let $\mathbf{b}_1, \ldots, \mathbf{b}_m$ and U be as in the theorem. Then we can find vectors $\mathbf{b}_{m+1}, \ldots, \mathbf{b}_r$ in U such that $\mathbf{b}_1, \ldots, \mathbf{b}_r$ is a base. Let B be the matrix with $\mathbf{b}_i$ as ith row, and let M consist of the m leftmost columns of B^{-1}. Then M is an $r \times m$ matrix of rank m that satisfies (3). □

We use Theorem 4.1 in order to determine under which conditions a set that is contained in the complement of a hyperplane is generic (r, m)-erasure correcting. The result is as follows.

Theorem 4.2. *Let $3 \leqslant m \leqslant r$. Let H be a hyperplane in $\mathbb{F}_2^r$, let $\mathbf{x} \notin H$, and let $K \subset H$. The set $\mathcal{A} := \mathbf{x} + (H \setminus K)$ is generic (r, m)-erasure correcting if and only if K does not contain a coset of an $(r - m)$-dimensional subspace of H.*

Proof. For notational convenience, we write $\mathcal{B} = \mathbb{F}_2^r \setminus \mathcal{A}$; note that

$$\mathcal{B} = H \cup (\mathbf{x} + K).$$

First, let U be an $(r - m)$-dimensional subspace of H, and assume that for some $\mathbf{c}$, the coset $\mathbf{c} + U$ is contained in K. Then U has a base $\mathbf{u}_1, \ldots, \mathbf{u}_{r-m}$ in H, and this base can be extended to a base $\mathbf{u}_1, \ldots, \mathbf{u}_{r-m}, \mathbf{b}_1, \ldots, \mathbf{b}_m$ with $\mathbf{b}_1, \ldots, \mathbf{b}_{m-1}$ in H and $\mathbf{b}_m = \mathbf{c} + \mathbf{x}$ not in H. Obviously all cosets $\mathbf{b}_i + U$ are now contained in $\mathcal{B}$: in fact in H for $i = 1, \ldots, m-1$, and in $\mathbf{x} + K$ for $i = m$. Theorem 4.1 implies that $\mathcal{A}$ is not (r, m)-erasure correcting.

Conversely, assume that $\mathcal{A}$ is not (r, m)-erasure correcting. According to Theorem 4.1, there exist an $(r - m)$-dimensional subspace U of $\mathbb{F}_2^r$ and m independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m$ for which $\operatorname{span}(\mathbf{b}_1, \ldots, \mathbf{b}_m) \cap U = \{\mathbf{0}\}$, such that $\mathcal{B} \supseteq \mathbf{b}_i + U$ for $i = 1, \ldots, m$. We investigate how this can happen, and distinguish between two cases.

Case (i). The subspace U is an $(r - m)$-dimensional subspace of H.

Note that a coset $\mathbf{b} + U$ of U is in H if and only if $\mathbf{b} \in H$. We conclude that there must be some $\mathbf{b}_i$ such that $\mathbf{b}_i \notin H$; consequently, $\mathbf{b}_i + U$ is outside H, and hence contained in $\mathbf{x} + K$; hence $\mathbf{b}_i + \mathbf{x} + U$ is contained in K.

Case (ii). The subspace U is not contained in H.

We write $U_0 := U \cap H$; note that U_0 is an $(r - m - 1)$-dimensional subspace of H. For $1 \leqslant i \leqslant m$, we define $\delta_i = 1$ if $\mathbf{b}_i \in H$, and $\delta_i = 0$ if $\mathbf{b}_i \notin H$. As

$$(\mathbf{b}_i + \delta_i \mathbf{x} + U_0) \cap H = \emptyset \quad \text{and} \quad \mathbf{b}_i + \delta_i \mathbf{x} + U_0 \subset \mathbf{b}_i + U \subseteq \mathcal{B} = H \cup (\mathbf{x} + K),$$

we have that

$$\mathbf{b}_i + \delta_i \mathbf{x} + U_0 \subseteq \mathbf{x} + K. \tag{4}$$

As $m \geqslant 3$, there are j and k, $1 \leqslant j < k \leqslant m$, and a $\delta \in \{0, 1\}$ such that $\delta_j = \delta_k = \delta$. It now follows from (4) that K contains $\mathbf{b}_j + (1 - \delta)\mathbf{x} + V$, where V is the $(r - m)$-dimensional subspace of H given by $V := \{\mathbf{0}, \mathbf{b}_j + \mathbf{b}_k\} + U_0$. □

The following characterization of special generic erasure-correcting sets will be used in Section 6.

Theorem 4.3. *Let H be a hyperplane in $\mathbb{F}_2^r$, and let $\mathcal{A} \subseteq \mathbb{F}_2^r \setminus H$. The set $\mathcal{A}$ is generic (r,m)-erasure correcting if and only if for all $(m-1)$ independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{m-1}$ in H and all $\epsilon_1, \ldots, \epsilon_{m-1} \in \mathbb{F}_2$, there is a vector $\mathbf{a} \in \mathcal{A}$ such that $(\mathbf{a}, \mathbf{v}_j) = \epsilon_j$ for $j = 1, \ldots, m-1$.*

Proof *(Sketch).* This is just a re-formulation of Theorem 4.2. Indeed, note that a coset X of an $(r-m)$-dimensional subspace C of an $(r-1)$-dimensional space can be described by $(r-1)-(r-m) = m-1$ independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{m-1}$ in $C^\perp$ together with $m-1$ values of the inner products of $\mathbf{v}_1, \ldots, \mathbf{v}_{m-1}$ with a vector in X. □

5. Constructions of special generic erasure correcting sets

In this section, we apply Theorem 4.2 to obtain explicit generic (r,m)-erasure correcting sets. We first give a construction for general r and m that slightly improves a construction in [5]. (The construction in [5] for the special case $m = 3$ has also been obtained, up to a linear transformation, by Weber and Abdel-Ghaffar in [2]; a slightly weaker result is given in [4].) Next, we will consider the special case $m = r-2$.

Our improvement depends on the following simple result.

Lemma 5.1. *Every coset of an $[n,k]$ code C contains a word of weight at most $n-k$; if $k \geqslant 2$ and $n-k \geqslant 2$, then every coset even contains a* non-zero *word of weight at most $n-k$.*

Proof. We may assume without loss of generality that C has generator matrix $G = (I|P)$, where I is the $k \times k$ identity matrix and P a binary $k \times (n-k)$ matrix. Consider a coset $\mathbf{x} + C$. Write $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$. Now C contains the word $(\mathbf{x}_1, \mathbf{x}_1 P)$, hence $\mathbf{x} + C$ contains $(\mathbf{0}, \mathbf{x}_2 + \mathbf{x}_1 P)$, which has weight at most $n-k$ and is non-zero except when the coset $\mathbf{x} + C$ is the code C itself.

Now, for $1 \leqslant i \leqslant k$ the code C contains the non-zero word $(\mathbf{e}_i \mid \mathbf{e}_i P)$, which has weight at most $n-k$ unless $\mathbf{e}_i P = \mathbf{1}$. Hence either one of these k words has weight at most $n-k$, or the code C contains the word $(\mathbf{e}_1 + \mathbf{e}_2, (\mathbf{e}_1 + \mathbf{e}_2)P) = (\mathbf{e}_1 + \mathbf{e}_2, \mathbf{0})$ of weight two. □

Theorem 5.2. *Let $1 \leqslant m \leqslant r$. The set $\mathcal{A}_{r,m}$, defined as*

$$\mathcal{A}_{r,m} = \left\{ \mathbf{x} = (x_1, x_2, \ldots, x_r) \in F_2^r \mid x_1 = 1 \text{ and } \mathrm{wt}(\mathbf{x}) \leqslant m \right\}$$

is generic (r,m)-erasure correcting, and has size $\sum_{i=0}^{m-1} \binom{r-1}{i}$.

If $r \geqslant m+2 \geqslant 5$, then the set $\mathcal{A}_{r,m} \setminus \{\mathbf{e}_1\}$ is generic (r,m)-erasure correcting as well.

Proof. Obviously, $\mathcal{A}_{r,m}$ has the size as claimed. Now, let $H = \{(x_1, \ldots, x_r) \in \mathbb{F}_2^r \mid x_1 = 0\}$, and let $K = \{\mathbf{y} \in H \mid \mathrm{wt}(\mathbf{y}) \geqslant m\}$. By Lemma 5.1, with $n = r-1$ and $k = r-m$, each coset of a subspace of H of dimension $k = r-m$ contains a word of weight at most $n-k = m-1$, and even such a non-zero word if $r-m \geqslant 2$ and $m-1 \geqslant 2$. So we conclude that neither the set K, nor the set $K \cup \{\mathbf{0}\}$ if $r-m \geqslant 2$ and $m-1 \geqslant 2$, contains a coset of an $(r-m)$-dimensional subspace of H. Since $\mathcal{A}_{r,m} = \mathbf{e}_1 + (H \setminus K)$, the result now follows from Theorem 4.2. □

Finding a special generic $(r, r-2)$-erasure correcting set of minimal size is equivalent to finding a longest code of codimension $r-1$ with minimum distance at least 5. The precise result is stated in the following proposition.

Proposition 5.3. *Let $r \geqslant 5$, and put $H := \{(x_1, x_2, \ldots, x_r) \in \mathbb{F}_2^r \mid x_1 = 0\}$. Let $\mathbf{0} \in K \subseteq H$ and write $M(K)$ to denote the $(r-1) \times (|K|-1)$ matrix consisting of the non-zero vectors of K*

with their first bit deleted. The set $\mathbf{e}_1 + (H \setminus K)$ *is generic* $(r, r-2)$*-erasure correcting if and only if the code with parity check matrix* $M(K)$ *has minimum distance at least five.*

Proof. It follows from Theorem 4.2 that $\mathbf{e}_1 + (H \setminus K)$ is generic $(r, r-2)$-erasure correcting if and only if K does not contain a coset of a 2-dimensional subspace of H.

The matrix $M(K)$ contains distinct non-zero columns, and so the code with parity check matrix $M(K)$ has minimum distance at least 3. Now, it is easy to check that four distinct vectors constitute a coset of a 2-dimensional space if and only if they sum to $\mathbf{0}$. Therefore, three columns of $M(K)$ add to zero if and only if these three vectors together with $\mathbf{0}$ form a 2-dimensional subspace of H. Similarly, four columns of $M(K)$ add to $\mathbf{0}$ if and only if they form a coset of a 2-dimensional subspace of H. □

Corollary 5.4. *Let* M *be a parity check matrix of an* $[n, n-r+1, 5]$ *code. Let* $K \subseteq \mathbb{F}_2^r$ *be defined as*

$$K = \mathbf{e}_1 \cup \left\{(1, x_2, \ldots, x_r) \in \mathbb{F}_2^r \mid (x_2, \ldots, x_r)^\top \text{ is a column of } M\right\}.$$

Then $\{(x_1, \ldots, x_r) \in \mathbb{F}_2^r \mid x_1 = 1\} \setminus K$ *is a generic* $(r, r-2)$*-erasure correcting set of size* $2^{r-1} - (n+1)$.

Example. In Corollary 5.4 we take $r = 2s+1$, and take for M the parity check matrix of an $[2^s - 1, 2^s - 2s - 1, 5]$ BCH code. In this way, we obtain a generic $(2s+1, 2s-1)$-erasure correcting set of cardinality $2^{2s} - 2^s$. Note that the set $\mathcal{A}_{2s+1,2s-1}$ from Theorem 5.2 (see also [5]) has cardinality $\sum_{i=0}^{2s-2} \binom{2s}{i} = 2^{2s} - \binom{2s}{2s} - \binom{2s}{2s-1} = 2^{2s} - 2s - 1$. We conclude that the construction based on the code with minimum distance 5 results in a generic $(2s+1, 2s-1)$ erasure correcting set that is much smaller than $\mathcal{A}_{2s+1,2s-1}$ for large s.

6. Definition, properties, and applications of (n, s)-good sets

In this section, we investigate special generic (r, m)-erasure correcting sets, based on the characterization of Theorem 4.3. In the first subsection, we define the notion of (n, s)-good sets and the closely related notion of (n, s)-**1** good sets. The relationship between these two notions is made explicit. In the second subsection, we use (n, s)-**1** good sets to explicitly construct generic $(r, 3)$-erasure correcting sets of a size about $3r^{\log_2 3}$—not linear in r, but much smaller than the size of $\mathcal{A}_{r,3}$ (which is about $\frac{1}{2}r^2$). The final subsection contains upper and lower bounds on the size of (n, s)-good sets.

6.1. (n, s)-good sets, (n, s)-**1** good sets, and their relation

We start with the following definition.

Definition 6.1. Let $1 \leqslant s \leqslant n$. A set $\mathcal{C} \subseteq \mathbb{F}_2^n$ is called (n, s)-*good* if for any s independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_s$ in $\mathbb{F}_2^n$ and for any $\epsilon_1, \ldots, \epsilon_s$ in $\mathbb{F}_2$, there is a $\mathbf{c} \in \mathcal{C}$ such that $(\mathbf{c}, \mathbf{v}_j) = \epsilon_j$ for $j = 1, \ldots, s$. The smallest size of an (n, s)-good set is denoted by $G(n, s)$.

Combination of Theorem 4.3 (with $H = \{(x_1, \ldots, x_r) \in \mathbb{F}_2^r \mid x_1 = 0\}$) and Definition 6.1 immediately yields the following result.

Proposition 6.2. *Let $2 \leqslant m \leqslant r$. A set $\mathcal{C} \subseteq \mathbb{F}_2^{r-1}$ is $(r-1, m-1)$-good if and only if $\{(1, \mathbf{c}) \mid \mathbf{c} \in \mathcal{C}\} \subseteq \mathbb{F}_2^r$ is generic (r, m)-erasure correcting.*

As a consequence, $G(r-1, m-1)$ is the smallest size of a special generic (r, m)-erasure correcting set, and hence $F(r, m) \leqslant G(r-1, m-1)$.

In view of Proposition 6.2, we aim to construct (n, s)-good sets, and to find bounds on $G(n, s)$. The following notion, closely related to that of (n, s)-good sets, turns out to be useful.

Definition 6.3. Let $1 \leqslant s \leqslant n$. A set $\mathcal{C} \subseteq \mathbb{F}_2^n$ is called (n, s)-**1** *good* if for any s independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_s$ in $\mathbb{F}_2^n$ there is a $\mathbf{c} \in \mathcal{C}$ such that $(\mathbf{c}, \mathbf{v}_j) = 1$ for $j = 1, \ldots, s$. The smallest size of an (n, s)-good set is denoted by $G_{\mathbf{1}}(n, s)$.

Remark. Note that a set $\mathcal{C} \subseteq \mathbb{F}_2^n$ is generic (n, s)-erasure correcting, (n, s)-good, or (n, s)-**1** good, if and only if for each $n \times s$ matrix V of rank s the image $V(\mathcal{C}) = \{\mathbf{c}V \mid c \in \mathcal{C}\}$ of the set $\mathcal{C}$ contains a vector of weight one, contains all vectors in $\mathbb{F}_2^s$, or contains the all-one vector, respectively.

It is clear that an (n, s)-good set is an (n, s)-**1** good set. The following lemma shows that the converse is "nearly" true.

Lemma 6.4. *Let $\mathcal{D}$ be an (n, s)-**1** good set. Then $C := \{\mathbf{0}\} \cup \mathcal{D}$ is an (n, s)-good set.*

Proof. Let $\mathbf{v}_1, \ldots, \mathbf{v}_s$ be independent vectors in $\mathbb{F}_2^n$, and let $\epsilon_1, \ldots, \epsilon_s$ be in $\mathbb{F}_2$. We will show that there is a $\mathbf{c} \in \mathcal{C}$ such that $(\mathbf{c}, \mathbf{v}_j) = \epsilon_j$ for $j = 1, 2, \ldots, s$.

If $\epsilon_j = 0$ for all j, then we take $\mathbf{c} = \mathbf{0}$. Otherwise, let i be such that $\epsilon_i = 1$. For $j = 1, 2, \ldots, s$, we define

$$\mathbf{w}_j = \mathbf{v}_j + (1 - \epsilon_j)\mathbf{v}_i.$$

Since $\mathbf{w}_i = \mathbf{v}_i$, we obviously have that $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_s) = \operatorname{span}(\mathbf{w}_1, \ldots, \mathbf{w}_s)$, hence the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_s$ are again independent.

As $\mathcal{D}$ is (n, s)-**1** good, there is a $\mathbf{c} \in \mathcal{D}$ such that $(\mathbf{c}, \mathbf{w}_j) = 1$ for all j; since $\mathbf{v}_i = \mathbf{w}_i$, it follows that $(\mathbf{c}, \mathbf{v}_j) = (\mathbf{c}, \mathbf{w}_j) + (1 - \epsilon_j)(\mathbf{c}, \mathbf{w}_i) = 1 + (1 - \epsilon_j) = \epsilon_j$. □

Lemma 6.5. *For $1 \leqslant s \leqslant n$, $G(n, s) = G_{\mathbf{1}}(n, s) + 1$.*

Proof. As we have seen in Lemma 6.4, the set $\{\mathbf{0}\} \cup \mathcal{D}$ is (n, s)-good whenever $\mathcal{D}$ is (n, s)-**1** good. This shows that $G(n, s) \leqslant G_{\mathbf{1}}(n, s) + 1$.

Next, we show that $G_{\mathbf{1}}(n, s) \leqslant G(n, s) - 1$. Let $\mathcal{C}$ be an (n, s)-good set of size $G(n, s)$, and let $\mathbf{x}$ be an arbitrary element from $\mathcal{C}$. We claim that $\mathbf{x} + \mathcal{C}$ is also (n, s)-good. Indeed, let $\mathbf{v}_1, \ldots, \mathbf{v}_s$ be s independent vectors in $\mathbb{F}_2^n$, and let $\epsilon_1, \ldots, \epsilon_s$ be in $\mathbb{F}_2$. Since $\mathcal{C}$ is (n, s)-good, there is a $\mathbf{c} \in \mathcal{C}$ such that $(\mathbf{c}, \mathbf{v}_j) = \epsilon_j + (\mathbf{x}, \mathbf{v}_j)$ for $j = 1, \ldots, s$; hence for this $\mathbf{c}$, we have $(\mathbf{x} + \mathbf{c}, \mathbf{v}_j) = \epsilon_j$ for all j. Clearly, $(\mathbf{x} + \mathcal{C}) \setminus \{\mathbf{0}\}$ is an (n, s)-**1** good set as well; its size equals $G(n, s) - 1$. □

6.2. Explicit construction of generic $(r, 3)$-erasure correcting sets of small size

Here we will give an explicit, recursive construction of $(r-1, 2)$-**1** good sets of cardinality at most $3(r-1)^{\log_2 3}$. By adding $\mathbf{0}$ to such a set, we obtain an $(r-1, 2)$-good set (cf. Lemma 6.4) and hence a generic $(r, 3)$-erasure correcting set with cardinality at most $3(r-1)r^{\log_2 3} + 1$.

The recursion step of the construction is described in the following theorem.

Theorem 6.6. *Let $2 \leqslant n \leqslant m$. Suppose that $\mathcal{C}_1$ is an $(n,2)$-**1** good set and that $\mathcal{C}_2$ is an $(m,2)$-**1** good set. Define $\mathcal{D} \subseteq \mathbb{F}_2^{n+m}$ as*

$$\big\{(\mathbf{c_1},\mathbf{0}) \mid \mathbf{c} \in \mathcal{C}_1\big\} \cup \big\{(\mathbf{0},\mathbf{c_2}) \mid \mathbf{c} \in \mathcal{C}_2\big\} \cup \big\{\big(s(\mathbf{c}),\mathbf{c}\big) \mid \mathbf{c} \in \mathcal{C}_2\big\},$$

*where $s(\mathbf{c})$ denotes the vector consisting of the first n coordinates of $\mathbf{c}$. Then $\mathcal{D}$ is an $(n+m,2)$-**1** good set, of size at most $|\mathcal{C}_1| + 2|\mathcal{C}_2|$.*

Proof. Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ and $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$ be two non-zero vectors in $\mathbb{F}_2^{n+m}$, with $\mathbf{x}_1$ and $\mathbf{y}_1$ in $\mathbb{F}_2^n$, and $\mathbf{x}_2$ and $\mathbf{y}_2$ in $\mathbb{F}_2^m$. We distinguish four cases.

(i) Both $\mathbf{x}_1$ and $\mathbf{y}_1$ are non-zero. As $\mathcal{C}_1$ is $(n,2)$-**1** good, it contains a vector $\mathbf{c}$ such that $(\mathbf{c},\mathbf{x}_1) = (\mathbf{c},\mathbf{y}_1) = 1$. Then the vector $\mathbf{d} = (\mathbf{c},\mathbf{0})$ from $\mathcal{D}$ satisfies $(\mathbf{d},\mathbf{x}) = (\mathbf{d},\mathbf{y}) = 1$.

(ii) Both $\mathbf{x}_2$ and $\mathbf{y}_2$ are non-zero. By a similar reasoning as in (i) we now find a vector $\mathbf{d} = (\mathbf{0},\mathbf{c}) \in \mathcal{D}$ that works.

(iii) We have that $\mathbf{x}_1 = \mathbf{0}$, $\mathbf{y}_2 = \mathbf{0}$ and both $\mathbf{x}_2, \mathbf{y}_1$ are non-zero. As $\mathcal{C}_2$ is $(m,2)$-**1** good, there is a vector $\mathbf{c} \in \mathcal{C}_2$ such that $(\mathbf{c},\mathbf{x}_2) = (\mathbf{c},(\mathbf{y}_1|\mathbf{0})) = 1$; as a consequence, the vector $\mathbf{d} = (s(\mathbf{c}),\mathbf{c})$ from $\mathcal{D}$ satisfies $(\mathbf{d},\mathbf{x}) = (\mathbf{d},\mathbf{y}) = 1$.

(iv) Finally, we have that $\mathbf{x}_2 = \mathbf{0}$, $\mathbf{y}_1 = \mathbf{0}$, and both $\mathbf{x}_1, \mathbf{y}_2$ are non-zero. By a similar reasoning as in (iii) we again find a vector $\mathbf{d} = (s(\mathbf{c}),\mathbf{c}) \in \mathcal{D}$ that works. □

Theorem 6.6 can be used to recursively construct $(n,2)$-**1** good sets. To get the construction started, we can use the $(2,2)$-**1** and $(3,2)$-**1** good sets

$$\{10, 01, 11\} \quad \text{and} \quad \big\{\mathbf{x} \in \mathbb{F}_2^3 \mid 1 \leqslant \mathrm{wt}(\mathbf{x}) \leqslant 2\big\}.$$

By repeated application of Theorem 6.6, starting with one of the above sets, we have the following corollary.

Corollary 6.7. *For all $m \geqslant 1$ we can construct a $(2^m,2)$-**1** good set of size at most 3^m, and a $(3 \cdot 2^{m-1},2)$-**1** good set of size at most $2 \cdot 3^m$.*

Puncturing an (n,s)-**1** good set in $(n-m)$ positions yields an (m,s)-**1** good set. That is, if $\mathcal{C} \subseteq \mathbb{F}_2^n$ is (n,s)-**1** good, then for each $m \leqslant n$, the set

$$\mathcal{D} = \big\{(x_1,\ldots,x_m) \in \mathbb{F}_2^m \mid \text{ there exist } x_{m+1},\ldots,x_n \text{ such that } (x_1,\ldots,x_n) \in \mathcal{C}\big\}$$

is (m,s)-**1** good. Indeed, let $\mathbf{v}_1,\ldots,\mathbf{v}_s$ be independent vectors in $\mathbb{F}_2^m$. For $1 \leqslant i \leqslant s$, let $\mathbf{w}_i \in \mathbb{F}_2^n$ be the vector $(\mathbf{v}_i,\mathbf{0})$. Clearly, the vectors $\mathbf{w}_1,\ldots,\mathbf{w}_s$ are independent. Now, there is a vector $\mathbf{d}$ that has inner product 1 with all $\mathbf{w}_i$'s. The vector consisting of the m leftmost entries of $\mathbf{c}$ is in $\mathcal{D}$, and has inner product 1 with all $\mathbf{v}_i$'s. Combining this observation with Corollary 6.7 yields the following result.

Corollary 6.8. *For all $n \geqslant 2$, we can construct an $(n,2)$-**1** good set of cardinality at most $3 \cdot n^{\log_2 3}$.*

Proof. Let $n \geqslant 2$, and let $m = \lceil \log_2 n \rceil$. We can construct an $(2^m,2)$-**1** good set of size 3^m, that can be punctured to an $(n,2)$-**1** good set of size at most $3^m < 3^{1+\log_2 n} = 3 \cdot n^{\log_2 3}$. □

Combination of Proposition 6.2, Lemma 6.4, and Corollary 6.8 immediately yields the following.

Corollary 6.9. *For all $r \geqslant 3$ we can construct a generic $(r, 3)$-erasure correcting set $\mathcal{A} \subseteq \mathbb{F}_2^r$ of size at most $1 + 3 \cdot (r-1)^{\log_2 3}$.*

6.3. Lower and upper bounds on the size of (n, s)-good sets

Here we will provide lower and upper bounds on $G(n, s)$. In particular, we will show (non-constructively) that $G(n, s)$ grows linearly in n for fixed s. We start with some lower bounds.

Lemma 6.10. *Let $1 \leqslant s \leqslant n$. If $\mathcal{C} \subseteq \mathbb{F}_2^n$ is (n, s)-**1** good, then $\mathcal{C}$ spans $\mathbb{F}_2^n$. As a consequence, $G_{\mathbf{1}}(n, s) \geqslant n$.*

Proof. If $\mathrm{span}(\mathcal{C}) \neq \mathbb{F}_2^n$, then $\mathrm{span}(\mathcal{C})^{\perp}$ contains a non-zero vector. This vector obviously has inner product 0 with all vectors in $\mathcal{C}$. □

Lemma 6.11. *For $n \geqslant 1$, we have that $G_{\mathbf{1}}(n, 1) = n$.*

Proof. As every non-zero vector in $\mathbb{F}_2^n$ has inner product 1 with at least one unit vector, the unit vectors in $\mathbb{F}_2^n$ constitute an $(n, 1)$-**1** good set of size n, and so $G_{\mathbf{1}}(n, 1) \leqslant n$. Now apply Lemma 6.10. □

Lemma 6.12. *If $2 \leqslant s \leqslant n$, then $G(n, s) \geqslant 2G(n-1, s-1)$.*

Proof. Let C be an (n, s)-good set. For $\epsilon \in \{0, 1\}$, we define $C_\epsilon \subseteq \mathbb{F}_2^{n-1}$ as

$$C_\epsilon = \big\{(x_1, \ldots, x_{n-1}) \in \mathbb{F}_2^{n-1} \mid (x_1, \ldots, x_{n-1}, \epsilon) \in C\big\}.$$

We will show that C_0 and C_1 both are $(n-1, s-1)$-good sets; as $|C| = |C_0| + |C_1|$, this implies the claim of the lemma.

Let $\epsilon \in \{0, 1\}$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_{s-1}$ be $s-1$ independent vectors in $\mathbb{F}_2^{n-1}$, and let $\epsilon_1, \ldots, \epsilon_{s-1}$ be in $\mathbb{F}_2$. For $1 \leqslant i \leqslant s-1$, we define $\mathbf{w}_i \in \mathbb{F}_2^n$ as $\mathbf{w}_i = (\mathbf{v}_i, 0)$; moreover, we define $\mathbf{w}_s := \mathbf{e}_n$. As C is (n, s)-good, there is a $\mathbf{c} \in C$ such that $(\mathbf{c}, \mathbf{w}_i) = \epsilon_i$ and $(\mathbf{c}, \mathbf{e}_n) = \epsilon$ for $i = 1, \ldots, s-1$. As a consequence, we can write $\mathbf{c} = (\mathbf{x}, \epsilon)$, with $\mathbf{x} \in C_\epsilon$. For $1 \leqslant i \leqslant s-1$, the vector $\mathbf{w}_i$ ends in a zero, and so $(\mathbf{x}, \mathbf{v}_i) = (\mathbf{c}, \mathbf{w}_i) = \epsilon_i$. □

Corollary 6.13. *For $1 \leqslant s \leqslant n$, we have that $G(n, s) \geqslant 2^{s-1}(n - s + 2)$.*

Proof. By induction on s, using Lemma 6.12 and the fact that $G(n, 1) = G_{\mathbf{1}}(n, 1) + 1 = n + 1$. □

Corollary 6.13 implies that for each fixed s, the function $G_{\mathbf{1}}(n, s)$ grows *at least* linearly in n, with coefficient at least 2^{s-1}. We will now show (non-constructively) that for each fixed s, the function $G_{\mathbf{1}}(n, s)$ in fact does not grow not faster than linear in n. More precisely, we will show the following.

Theorem 6.14. *Let $1 \leqslant s \leqslant n$. Then we have that*

$$G_{\mathbf{1}}(n,s) \leqslant d_s n - e_s, \quad \text{where } d_s = \frac{s}{-\log_2(1-2^{-s})} \text{ and } e_s = \frac{\log_2 s!}{-\log_2(1-2^{-s})}.$$

Proof. We again apply the probabilistic method. We define $\mathcal{Y}$ as the collection of all s-sets of independent vectors in $\mathbb{F}_2^n$. Note that obviously

$$|\mathcal{Y}| < \frac{1}{s!}2^{ns}.$$

We randomly pick an N-subset A of $\mathbb{F}_2^n$. For each $Y \in \mathcal{Y}$, we define the random variable X_Y as

$$X_Y = \begin{cases} 0, & \text{if there is an } \mathbf{a} \in A \text{ such that for all } \mathbf{y} \in Y,\ (\mathbf{a},\mathbf{y}) = 1; \\ 1, & \text{otherwise.} \end{cases}$$

Note that A is (n,s)-$\mathbf{1}$ good if $\sum_{Y\in\mathcal{Y}} X_Y < 1$. Consequently, there is an N-subset A of $\mathbb{F}_2^n$ whenever $E[\sum_{Y\in\mathcal{Y}} X_Y] < 1$.

For each $Y \in \mathcal{Y}$, there are 2^{n-s} vectors $\mathbf{x}$ in $\mathbb{F}_2^n$ such that $(\mathbf{x},\mathbf{y}) = 1$ for all $\mathbf{y} \in Y$ (these are precisely the vectors in a coset of $(\mathrm{span}(Y))^{\perp}$). As a consequence, $E[X_Y] = \binom{2^n - 2^{n-s}}{N} \Big/ \binom{2^n}{N}$, and so

$$E\left[\sum_{Y\in\mathcal{Y}} X_Y\right] = \sum_{Y\in\mathcal{Y}} E[X_Y] = |\mathcal{Y}| \binom{2^n - 2^{n-s}}{N} \Big/ \binom{2^n}{N} < \frac{1}{s!}2^{ns}\left(1-2^{-s}\right)^N. \qquad \square$$

It is noteworthy that for each $m \geqslant 2$, the linear part in the upper bound on $F(r,m)$ implied by Theorem 6.14 exceeds the upper bound on $F(r,m)$ from Section 3. So for large r, the probabilistic construction from Section 3 provides a smaller result than the construction provided here.

7. Conclusions

In this paper we have introduced the notion of generic (r,m)-erasure correcting sets in $\mathbb{F}_2^r$; such sets provide for each binary code C of codimension r a collection of parity checks for C that can be used to iteratively correct all correctable erasure patterns of size at most m. Our main result is that the minimal size $F(r,m)$ of such sets is *linear in* r for fixed m. We provide various explicit constructions of generic (r,m)-erasure correcting sets, which in certain cases improve upon previous results for specific codes, notably for $m = 3$.

We have also introduced the related notion of an (r,m)-good subset, where a collection $\mathcal{C} \subseteq \mathbb{F}_2^r$ is (r,m)-good if, for every $r \times m$ matrix V of rank m, the image $V(\mathcal{C})$ of $\mathcal{C}$ equals $\mathbb{F}_2^m$ (for generic (r,m)-erasure correcting sets, we require that the image contains a vector of weight one). We showed that the minimal size $G(r,m)$ of an (r,m)-good set is linear in r for fixed m, and satisfies $F(r+1,m+1) \leqslant G(r,m)$.

The main remaining problem is to find *explicit* constructions for (r,m)-erasure correcting sets and $(r,m-1)$-good sets of size linear in r, especially in the first open case $m = 3$.

Acknowledgments

We thank both referees for their very careful reading of the manuscript.

References

[1] C. Di, D. Proietti, I.E. Telatar, T.J. Richardson, R.L. Urbanke, Finite-length analysis of low-density parity-check codes on the binary erasure channel, IEEE Trans. Inform. Theory 48 (6) (June 2002) 1570–1579.
[2] J.H. Weber, K.A.S. Abdel-Ghaffar, Stopping set analysis for hamming codes, in: M.J. Dinneen (Ed.), Proc. IEEE ITSOC Information Theory Workshop 2005 on Coding and Complexity, Rotorua, New Zealand, August 29–September 1, ISBN 7803-9481-X, 2005, pp. 244–247.
[3] M. Schwartz, A. Vardy, On the stopping distance and the stopping redundancy of codes, in: Proc. 2005 IEEE International Symposium on Information Theory, Adelaide, Australia, 4–9 September, 2005, pp. 975–979.
[4] M. Schwartz, A. Vardy, On the stopping distance and the stopping redundancy of codes, IEEE Trans. Inform. Theory 52 (3) (March 2006) 922–932.
[5] H.D.L. Hollmann, L.M.G.M. Tolhuizen, On parity check collections for iterative erasure decoding that correct all correctable erasure patterns of a given size, IEEE Trans. Inform. Theory, submitted for publication, see also arXiv: cs.IT/0507068.
[6] N. Alon, J. Spencer, The Probabilistic Method, Wiley–Interscience, New York, 1992.